Comprehensive Survival Analysis of Heart Transplant Outcomes Using the Stanford2 Dataset

Rupanjan Mukherjee[†], Sourav Biswas[‡], Subhendu Ghosh[††]

[†‡††]*Indian Statistical Institute, Kolkata*

Roll Numbers: [†] MD2314, [‡] MD2320, [††] MD2322

✉ E-mail: [†] rup629063@gmail.com, [‡] sourav3biswas2003@gmail.com, [††] subhendug051@gmail.com

Abstract

This study analyzes survival outcomes in heart transplant patients using the stanford2 dataset from R's survival package, incorporating time-to-event data, age, and T5 mismatch score. Employing non-parametric (Kaplan-Meier, Nelson-Aalen), hypothesis testing (Log-rank, Gehan, Tarone-Ware), semi-parametric (Cox proportional hazards), and parametric (Weibull, log-normal AFT) methods, we evaluate model performance via AIC and compare parametric curves to Kaplan-Meier estimates. The results aim to identify an optimal prognostic model to guide clinical management of end-stage heart failure patients.

Keywords: Accelerating Failure Time Model, Akaike Information Criterion, Breslow's Baseline Hazard Esimate, Cox-Snell Residuals, Cox Proportional Hazard Model, Equi-Probable Band, Generalized Gehan's Test, Greenwood's Formula, Hall-Wellner Band, Kaplan-Meier Survival Curve, Log-Rank Test, Maximum Partial Likelihood Estimate, Nelson-Aalen Cumulative Hazard function, Omnibus χ^2 Test, Schoenfeld Residuals, Tarone-Ware Test.

1 Introduction

Heart transplantation represents the definitive therapeutic intervention for a select cohort of patients diagnosed with end-stage heart failure. The precise modeling of survival times is paramount for clinicians, enabling informed prognosis and effective patient management. This investigation undertakes a thorough survival analysis utilizing the `stanford2` dataset, accessible within R's `survival` package, which encapsulates time-to-event data alongside pertinent covariates, namely age and T5 mismatch score.

To elucidate survival dynamics, this study employs an array of statistical methodologies. Non-parametric techniques, including the Kaplan-Meier and Nelson-Aalen estimators, are utilized to characterize survival distributions. Hypothesis testing is conducted through the application of the Log-rank, Gehan, and Tarone-Ware tests to assess differences across groups. Furthermore, the semi-parametric Cox proportional hazards model is implemented to evaluate covariate effects, complemented by parametric Accelerated Failure Time (AFT) models based

on Weibull and log-normal distributions. Model performance is rigorously compared using the Akaike Information Criterion (AIC), while the superposition of parametric survival curves against the Kaplan-Meier estimate serves to delineate the most efficacious prognostic framework.

This comprehensive analytical approach aims to enhance the understanding of survival outcomes in heart transplantation, thereby supporting clinical decision-making.

2 About the Data

The `stanford2` dataset provides the following variables for survival analysis:

- **`time`**: Survival time in days post heart transplantation.
- **`status`**: Event indicator (1 = death, 0 = censored).
- **`age`**: Patient's age at transplantation in years.
- **`t5`**: T5 mismatch score, a measure of donor-recipient incompatibility.

Source : Escobar, L. A., & Meeker, W. Q. Jr. (1992). Assessing influence in regression analysis with censored data. *Biometrics*, 48, 507–528. Page 519.

We have a total of $n = 184$ observations, of which 71 observations are right censored. Call the covariates age and T5 mismatch score by Z_1 and Z_2 respectively. For ith individual, survival time is t_i, age z_{1i}, t5 mismatch score z_{2i} and censoring status δ_i.

3 Research Objectives

1. Estimate the overall survival function for heart transplant patients in the Stanford2 dataset.
2. Assess survival differences across quartiles of T5 mismatch score and patient age.
3. Conduct hypothesis tests to compare survival distributions among stratified groups.
4. Develop and validate a Cox proportional hazards model to evaluate covariate effects.
5. Fit Weibull and log-normal Accelerated Failure Time (AFT) models and compare their performance using the Akaike Information Criterion (AIC).
6. Compare non-parametric and parametric survival curves to determine the optimal prognostic model.

4 Data Preparation

The `stanford2` dataset is loaded, and variables (`id`, `time`, `status`, `age`, `t5`) are selected. Missing values are excluded. Age and T5 mismatch score are divided into four quartile groups, labeled as `age_group` (Q1 to Q4) and `t5_group` (Low, Medium, High, Very High). The resulting dataset, `stan2`, is previewed in Table 1 showing the first six rows.

Table 1: Preview of stanford2 with Quartile Groups

id	time	status	age	t5	age_grou	t5_group
139	86	1	12	1.26	Q1	High
159	10	1	13	1.49	Q1	Very High
119	1116	0	14	0.54	Q1	Low
74	2006	0	15	1.26	Q1	High
120	1107	0	18	0.25	Q1	Low
99	1232	1	18	0.70	Q1	Medium

5 Kaplan-Meier Survival Curve

Say, $Y_j = \#$ individuals at risk at time t_j and $d_j = \#$ events at time t_j. Then the KM Estimator of the survival curve is given by,

$$\hat{S}_{KM}(t) = \prod_{j:t_j \leq t} \left(1 - \frac{d_j}{Y_j}\right),$$

with Greenwood's pointwise variance,

$$\widehat{\text{Var}}\left(\hat{S}(t)\right) = \left(\hat{S}(t)\right)^2 \sum_{j:\, t_j \leq t} \frac{d_j}{(Y_j - d_j)\, Y_j}.$$

Klein and Moeschberger [2003]. The survival curve displays the estimated survival probability over time (in days) for the entire cohort of heart transplant patients from the Stanford2 dataset. The x-axis represents time post-transplantation, and the y-axis shows the survival probability, which decreases as time progresses due to patient mortality. The curve provides a visual summary of the overall survival trend.

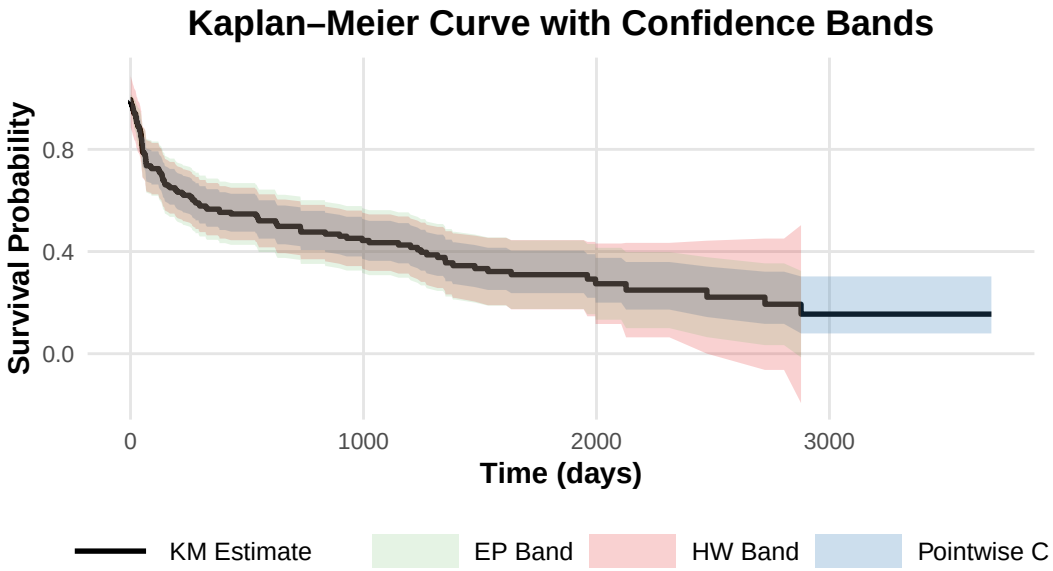

Figure 1: Overall Kaplan-Meier Survival Curve with Confidence Bands, Pointwise CI

5.1 Survival Estimates at Yearly Intervals

Table 2 below presents survival statistics at annual time points:

Table 2: KM Estimates at Yearly Intervals

Time	AtRisk	Events	Survival	Variance	Lower	Upper
0	157	0	1.0000	0.0000	1.0000	1.0000
365	84	69	0.5588	0.0016	0.4861	0.6424
730	62	11	0.4784	0.0017	0.4048	0.5655
1095	49	5	0.4362	0.0017	0.3622	0.5253
1460	30	8	0.3535	0.0018	0.2790	0.4480
1825	21	3	0.3173	0.0019	0.2430	0.4142
2190	10	3	0.2554	0.0023	0.1770	0.3686
2555	8	1	0.2270	0.0025	0.1472	0.3502
2920	4	2	0.1589	0.0029	0.0813	0.3105
3285	2	0	0.1589	0.0029	0.0813	0.3105
3650	1	0	0.1589	0.0029	0.0813	0.3105

5.1.1 Survival Analysis by T5 Mismatch Score Quartile

To investigate the relationship between T5 mismatch scores and survival outcomes, we generated Kaplan-Meier survival curves stratified by quartiles of the T5 mismatch score:

Low, Medium, High, and Very High. The survival plot in the following figure illustrates how survival probability changes over time for each group.

- **Low Mismatch Group**: This group exhibits the highest survival probabilities, with only a gradual decline over time.
- **Medium Mismatch Group**: The survival curve is slightly below that of the Low group but still shows relatively favorable survival outcomes.
- **High Mismatch Group**: A more pronounced decline in survival probability is observed, especially at later time points.
- **Very High Mismatch Group**: This group demonstrates the steepest drop in survival probability, suggesting poorer outcomes compared to the other quartiles.

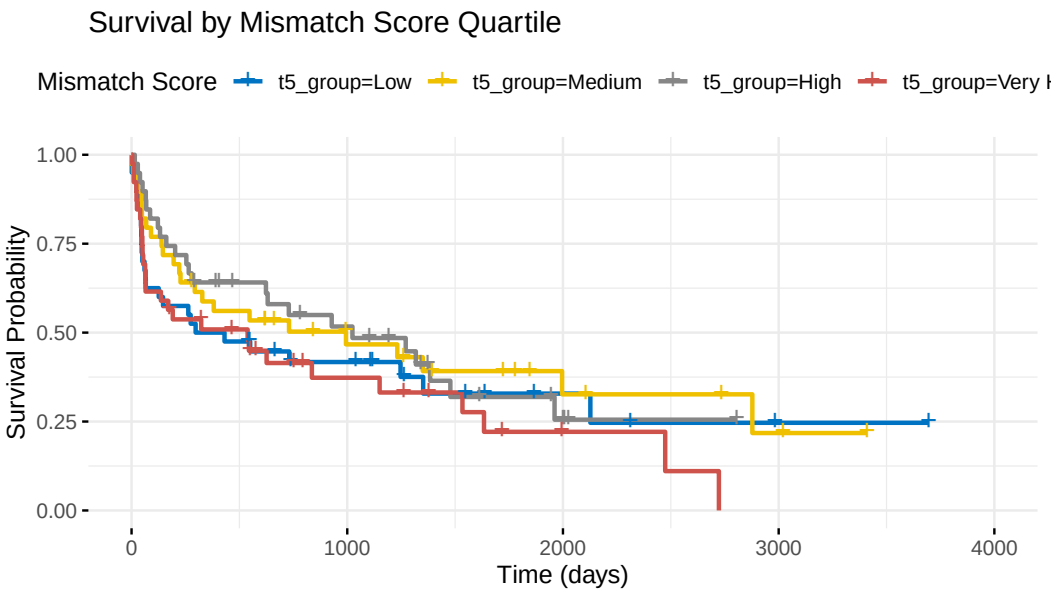

Figure 2: Survival Curves by T5 Mismatch Scores

These trends indicate a potential association between higher T5 mismatch scores and reduced survival. However, statistical significance would need to be confirmed with additional tests (e.g., log-rank test).

5.2 Survival Analysis by Age Quartile

The survival plot displays four Kaplan-Meier survival curves, one for each age quartile:

- **Q1 (Youngest)**: Typically shows the highest survival probability, with a slower decline over time, indicating better outcomes for younger patients.
- **Q2**: Exhibits a slightly steeper decline compared to Q1 but generally maintains favorable survival.
- **Q3**: Shows a more noticeable decrease in survival probability, suggesting moderate age-related impacts.
- **Q4 (Oldest)**: Demonstrates the steepest decline, indicating that older patients may have poorer survival outcomes.

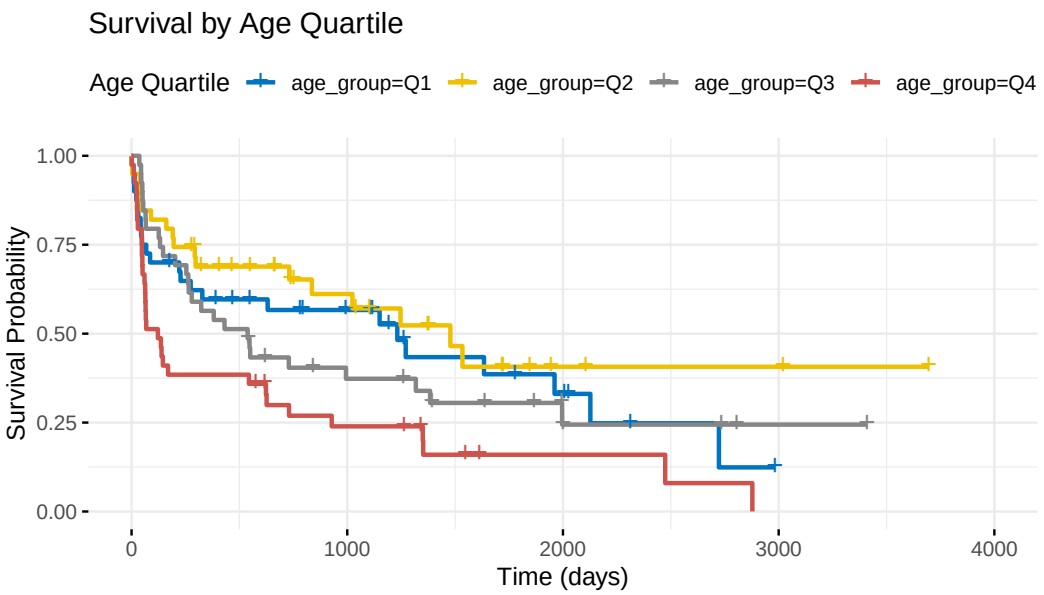

Figure 3: Survival Curves by Age

These trends suggest that older age at transplantation may be associated with reduced survival, though statistical significance requires further testing (e.g., log-rank test).

6 Nelson-Aalen Cumulative Hazard Estimation

The Nelson-Aalen Cumulative Hazard estimate is,

$$\tilde{H}(t) = \begin{cases} 0 & t < t_1 \\ \sum_{j:\, t_j \le t} \frac{d_j}{Y_j} & \text{o.w.} \end{cases}$$

In a similar manner we have estimated the Cumulative Hazard Function $H(t)$, overall and based on `age_group` and `t5_group`. The three plots are shown below.

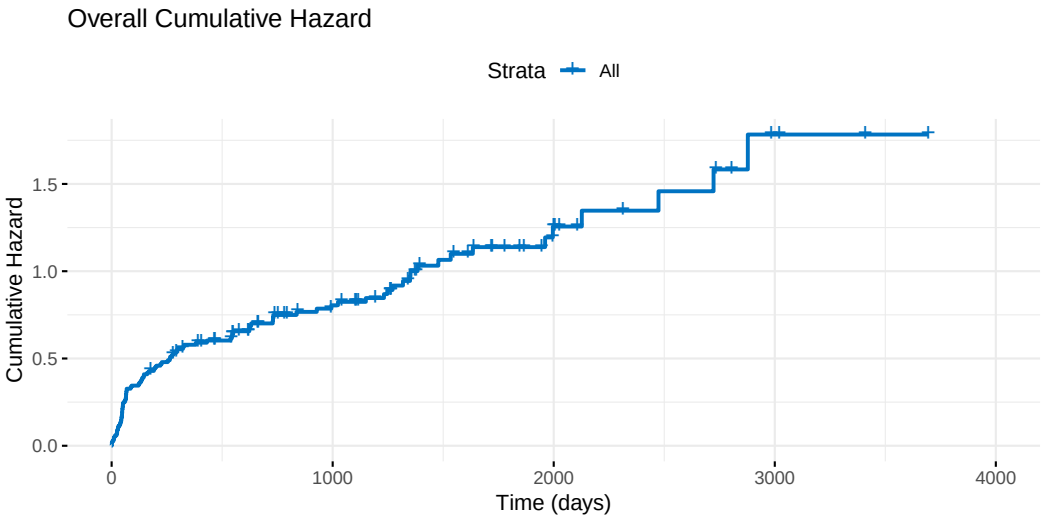

Figure 4: Nelson-Aalen Estimate of Cumulative Hazard Function

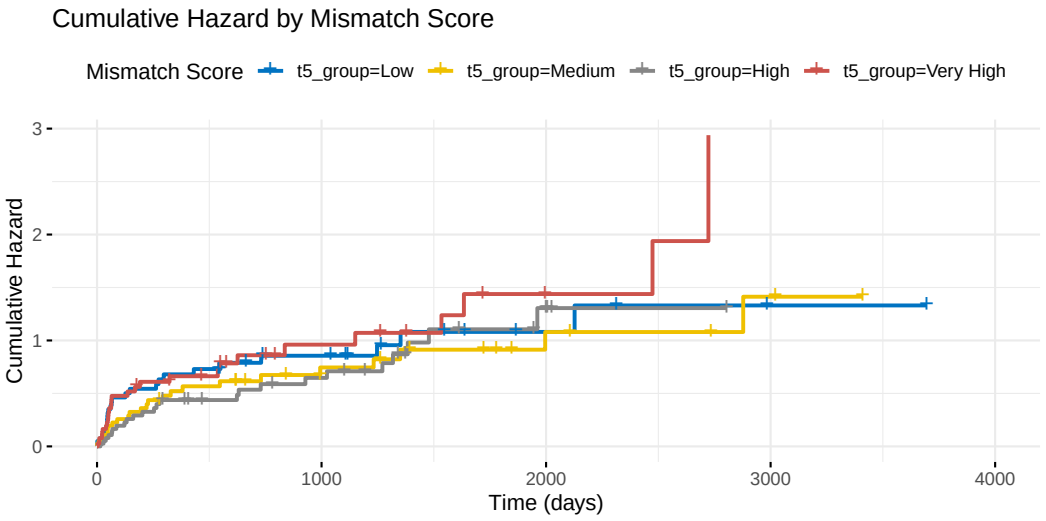

Figure 5: Cumulative Hazard by T5 mismatch scores

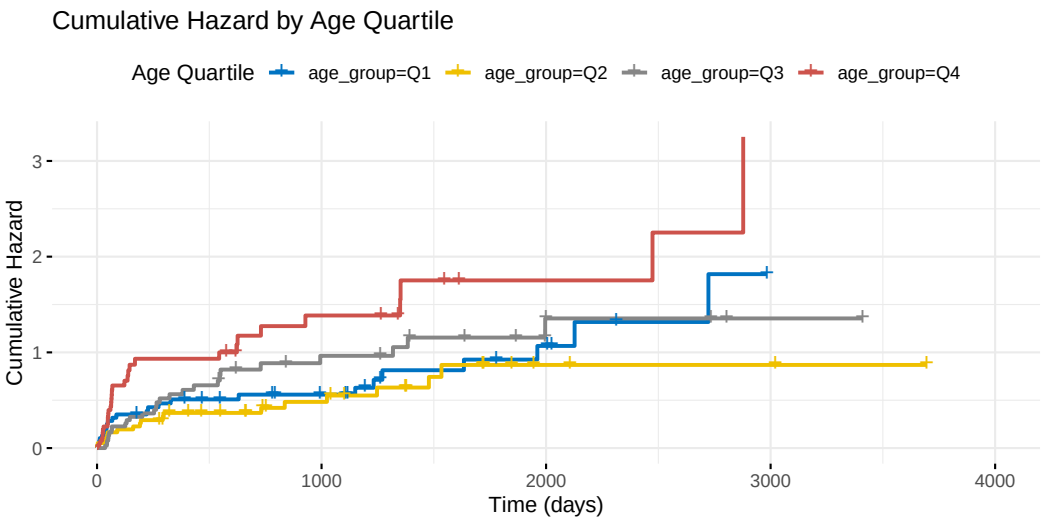

Figure 6: Cumulative Hazard by Age

7 Hypothesis Tests for Survival Differences

We have created 4 groups based on each coavariate, age (Z_1) and t5 score (Z_2). Suppose F_k, G_k be distribution functions of lifetime and censoring RV (C_i) for the four groups. We are about to test,

$$H_0^* : \{F_1 = F_2 = F_3 = F_4\} \cap \{G_1 = G_2 = G_3 = G_4\}$$

against $H_1^* : H_0^*$ is false. Survival distributions were compared across T5 mismatch score and age quartile groups using three hypothesis tests: Log-rank, Gehan-Wilcoxon, and Tarone-Ware. Each test was performed separately for `t5_group` and `age_group`. Test statistics (chi-square) and p-values were calculated to determine if survival differences are statistically significant at level $\alpha = 0.05$ ($p < 0.05$). Results are summarized in a table with grouped headers for each test type, showing test statistics and p-values for both covariates.

Table 3 presents the results of the hypothesis tests for survival differences across T5 mismatch score and age quartiles:

Table 3: Hypothesis Tests for Survival Differences

	Log-rank Test		Gehan's Test		Tarone–Ware Test	
Covariate	**TS**	**p-value**	**TS**	**p-value**	**TS**	**p-value**
Mismatch Score	3.18	0.365	4.08	0.253	3.56	0.313
Age Quartile	13.71	0.00332	13.17	0.00427	13.64	0.00344

7.1 Interpretation

- **Mismatch Score**: The Log-rank test yields a test statistic of 3.18 ($p = 0.365$), Gehan-Wilcoxon 4.08 ($p = 0.253$), and Tarone-Ware 3.56 ($p = 0.313$). All p-values are above 0.05, indicating no statistically significant differences in survival across T5 mismatch score quartiles. This suggests that the degree of donor-recipient mismatch, as measured by T5, may not strongly influence survival in this cohort.
- **Age Quartile**: The Log-rank test gives a test statistic of 13.71 ($p = 0.003$) , Gehan-Wilcoxon 13.17 ($p = 0.004$), and Tarone-Ware 13.64 ($p = 0.003$). All p-values are below 0.05, confirming significant differences in survival across age quartiles. Older patients (Q4) likely experience worse survival outcomes compared to younger patients (Q1), consistent with the Kaplan-Meier curves observed earlier.

Based on the hypothesis test results, the age quartile (Z_1) was identified as a significant covariate influencing survival outcomes ($p < 0.05$ across Log-rank, Gehan-Wilcoxon, and Tarone-Ware tests), whereas the T5 mismatch score showed no significant effect ($p > 0.05$).

Consequently, subsequent survival modeling, including the Cox proportional hazards and Accelerated Failure Time (AFT) models, will incorporate age quartile as the primary covariate to explore its impact on survival time. This approach ensures that the analysis focuses on the variable with statistically significant differences, providing a robust basis for understanding how age influences post-transplantation survival in the Stanford2 dataset.

8 Cox PH Model

The Cox PH model is the following,

$$h\left(t|\, Z_1 = z\right) = h_0\left(t\right) e^{\beta z},$$

where $h_0\left(\boldsymbol{.}\right)$ is the baseline hazard function.

8.1 Graphical Check

The Cox-Snell residuals (r_i) provides a quick overview of the applicability of Cox PH Model, given by,

$$r_i = \hat{H}_0\left(t\right) e^{z_{1i}\hat{\beta}},$$

with $\hat{H}_0$ is estimated H_0 from Breslow's Baseline Hazard and $\hat{\beta}$ is the Maximum Partial Likelihood estimate of β. Then $-\ln \hat{S}_{KM}\left(r_i\right)$ against r_i is plotted.

To assess the goodness-of-fit of the Cox proportional hazards model with age quartile as the covariate, Cox-Snell residuals were calculated and analyzed. These residuals, representing the estimated cumulative hazard for each observation, were used to construct a Kaplan-Meier survival curve. The cumulative hazard of the residuals was plotted against the residuals themselves. A well-fitted model should produce a plot that approximates a 45-degree line.

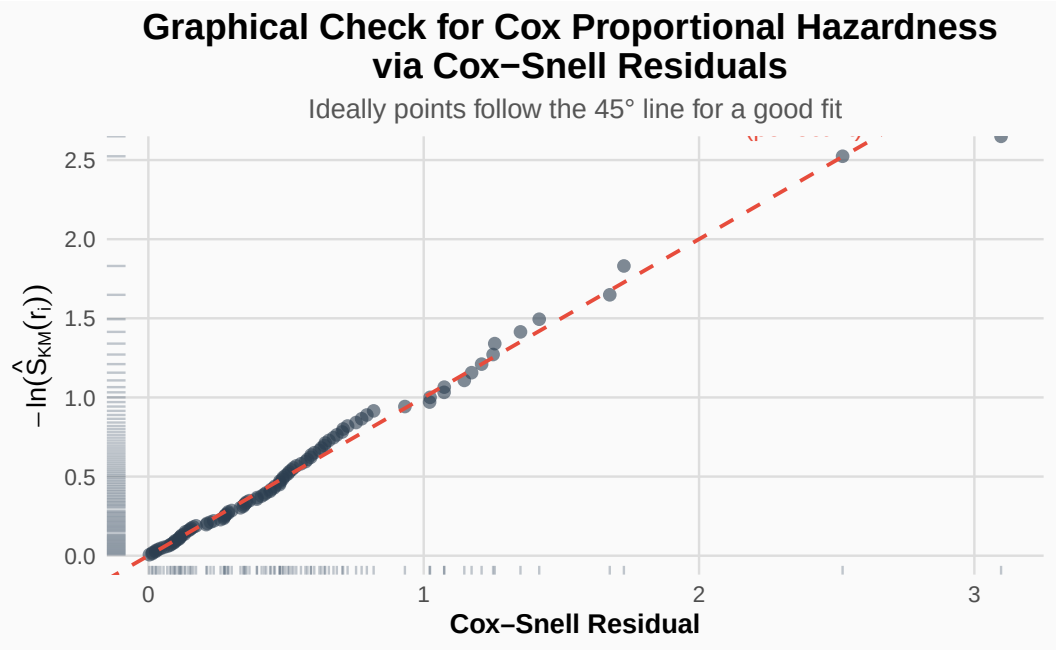

Figure 7: Cox-Snell residual plot

8.1.1 Interpretation

The close adherence of the Cox-Snell residuals to the 45-degree line suggests that the Cox PH model is well-specified for the Stanford2 dataset when using age quartile as the covariate. This confirms that the model effectively describes the relationship between age groups and survival outcomes. However, to ensure robustness, additional diagnostics, such as testing the proportional hazards assumption using Schoenfeld residuals, could further validate the model's assumptions.

8.2 PH assumption check

Table 4 presents the results of the proportional hazards assumption test for the Cox PH model:

Table 4: Proportional Hazards Assumption Test

Covariate	Chi-Square	DF	p-value
age	0.83	1	0.362
GLOBAL	2.77	2	0.251

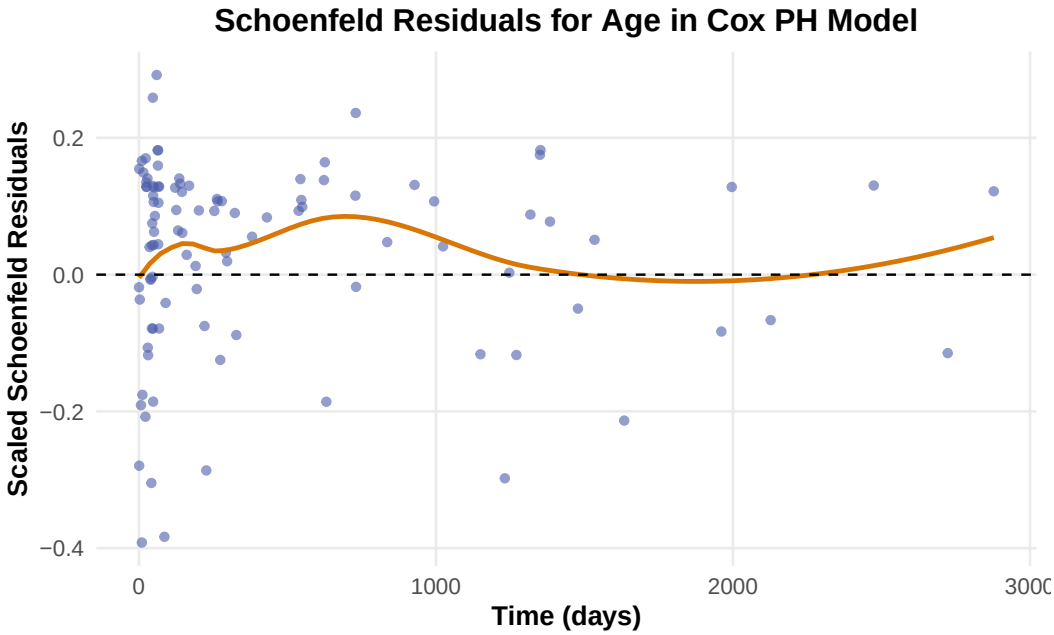

Figure 8: Schoenfeld Residuals for Age

The chi-square test for age yields a statistic of 0.83 with a p-value of 0.362 (Table 4), indicating no evidence of a time-varying effect ($p > 0.05$). The global test ($\chi^2 = 2.77, p = 0.251$) further supports the model's overall adherence to the PH assumption. The Schoenfeld residual plot reinforces this, showing no systematic trend, which confirms that the hazard ratio for age remains constant over time. These results validate the use of the Cox PH model with age as a covariate in the Stanford2 dataset.

8.3 Model fitting

Table 5 presents the estimates from the Cox PH model for the age covariate:

Table 5: Cox PH Model Estimates

Covariate	HR	X95..CI	p.value
age	1.03	[1.007, 1.053]	0.0091

9 AFT Model

An Accelerating Failure Time (AFT) model is of the form,

$$\ln T = \alpha + \beta Z_1 + \sigma W$$

where $\sigma > 0$ and $W \sim (0, 1)$ Kalbfleisch and Prentice [2002].

Two AFT models were fitted to the Stanford2 dataset using age as a continuous covariate: one with a *Weibull* error distribution and one with a *Lognormal* error distribution. The models estimate the acceleration factor for age, indicating how a one-year increase in age affects survival time. A table summarizes the coefficients, standard errors, p-values, and AIC for both models. A plot compares the predicted survival curves from both AFT models against the Kaplan-Meier (KM) estimate to visually assess model fit.

Table 6 presents the estimates from the Weibull and lognormal AFT models for the age covariate:

Table 6: AFT Model Estimates

Model	Covariate	Coefficient	SE	p-value	AIC
Weibull	Age	-0.057	0.021	0.005	1576.03
Lognormal	Age	-0.039	0.020	0.047	1575.25

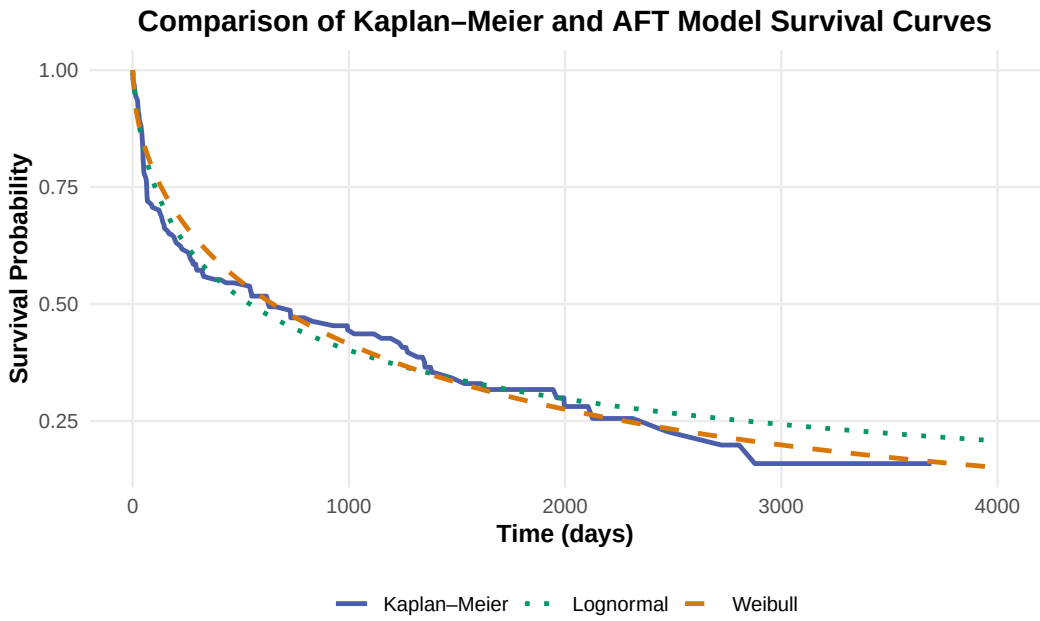

Figure 9: KM and AFT Model Survival Curves (manual AFT)

The above figure displays the Kaplan-Meier survival curve alongside the predicted survival curves from the Weibull and lognormal AFT models, evaluated at the mean age. Both AFT curves align closely with the KM curve, indicating good model fit. The lognormal model shows a slightly better fit, as evidenced by its lower AIC (1575.25 vs. 1576.03). The Weibull model slightly underestimates survival at early time points, while the lognormal model tracks the KM curve more consistently across the time range.

9.1 Interpretation

The Weibull AFT model estimates a coefficient of -0.057 for age ($p = 0.005$) , indicating that each one-year increase in age reduces the expected survival time by approximately 5.5% (Table 6). The lognormal AFT model estimates a coefficient of -0.039 ($p = 0.047$), suggesting a 3.8% reduction per year. Both models confirm that older age significantly shortens survival time, with the lognormal model's lower AIC (1575.25 vs. 1576.03) indicating a slightly better balance of fit and complexity. The survival curve plot supports this, showing that the lognormal model's predicted curve aligns more closely with the KM estimate. These results reinforce age as a key predictor of survival in the Stanford2 cohort, consistent with prior Cox PH model findings.

10 Conclusion

The survival analysis of the Stanford2 heart transplant dataset identifies patient age as a key predictor of survival, while the T5 mismatch score shows no significant impact. Non-parametric Kaplan-Meier and Nelson-Aalen estimators reveal poorer survival for older patients, confirmed by hypothesis tests showing significant differences across age groups but not T5 mismatch levels. The Cox proportional hazards model, validated by residual diagnostics, underscores age's role in increasing hazard rates. Parametric Weibull and lognormal AFT models quantify age's effect on survival time, with the lognormal model offering a slightly better fit. These findings emphasize age as a critical factor in heart transplant outcomes, guiding clinical strategies while downplaying the prognostic role of T5 mismatch.

11 Contribution

This project was a collaborative effort among three team members: Sourav Biswas, Subhendu Ghosh, and Rupanjan Mukherjee. Sourav was responsible for collecting and preparing the data, ensuring it was clean and suitable for analysis. He also formatted the final report, giving it a professional and polished appearance. Additionally, Sourav conducted hypothesis testing and fitted the Proportional Hazards (PH) model, performing rigorous statistical analysis to validate the findings. Subhendu focused on fitting the Accelerated Failure Time (AFT) model, selecting the appropriate distribution and estimating its parameters to provide valuable insights into the data. Rupanjan handled the hazard survival estimation, calculating survival probabilities and hazard rates, which were critical for understanding the survival dynamics in the dataset. Each team member wrote their own code for their respective contributions and authored the corresponding sections of the report, ensuring accuracy and clarity in the documentation. Throughout the project, all members actively participated, bringing their expertise and dedication to successfully achieve the project's objectives.

Entire analysis has been implemented in R. You can find the codes on GitHub.

References

John D. Kalbfleisch and Ross L. Prentice. *The Statistical Analysis of Failure Time Data*. Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken, NJ, 2nd edition, 2002. URL https://doi.org/10.1002/9781118032985.

John P. Klein and Melvin L. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Statistics for Biology and Health. Springer, New York, 2nd edition, 2003. URL https://doi.org/10.1007/b97377.